ESTIMATION OF OBESITY LEVELS BASED ON EATING HABITS AND PHYSICAL CONDITIONS

Abstract

This study aims to estimate obesity levels by examining the relationships between eating habits, physical activity, and the likelihood of obesity. A regression model is developed using a comprehensive dataset to predict obesity based on key variables.

Dataset

This dataset includes information for predicting obesity levels in individuals from Mexico, Peru, and Colombia, considering their eating habits and physical condition. It consists of 17 attributes and encompasses a total of 2111 records.

The dataset was taken from UCI Machine Learning Repository;

https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition

Methodology

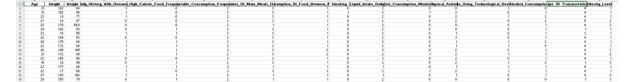
This study utilizes a dataset for estimating obesity levels in individuals from Mexico, Peru, and Colombia based on their eating habits and physical condition. The initial step involves cleaning and preprocessing the data, addressing aspects such as column names, data types, and converting categorical values to numerical ones. Subsequently, data visualization techniques are applied to uncover patterns and relationships within the dataset. Statistical methods, including the use of correlation matrices, are then employed to identify relevant features. To construct a regression model, the dataset is divided into a training set and a test set. Standard scaling is

implemented to normalize the data. Following model construction, an evaluation of its performance is conducted to assess its effectiveness in estimating obesity levels.

Before Preprocessing

3ender	Age	Height	Weight	family_histor	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
emale	21	1.62	64	yes	no	2		3 Sometimes	no		2 no		0	1 no	Public_Transportation	Normal_Weight
emale	21	1.52	56	yes	no	3		3 Sometimes	yes		3 yes		3	0 Sometimes	Public_Transportation	Normal_Weight
Viale	23	1.8	77	yes	no	2		3 Sometimes	no		2 no		2	1 Frequently	Public_Transportation	Normal_Weight
Viale	27	1.8	87	no	no	3		3 Sometimes	no		2 no		2	0 Frequently	Walking	Overweight_Level_I
Viale	22	1.78	89.8	no	no	2		1 Sometimes	no		2 no		0	0 Sometimes	Public_Transportation	Overweight_Level_II
Viale	29	1.62	53	no	yes	2		3 Sometimes	no		2 no		0	0 Sometimes	Automobile	Normal_Weight
emale	23	1.5	55	yes	yes	3		3 Sometimes	no	1	2 no		1	0 Sometimes	Motorbike	Normal_Weight
Vale	22	1.64	53	no	no	2		3 Sometimes	no		2 no		3	0 Sometimes	Public_Transportation	Normal_Weight
Vale	24	1.78	64	yes	yes	3		3 Sometimes	no		2 no		1	1 Frequently	Public_Transportation	Normal_Weight
Viale	22	1.72	68	yes	yes	2		3 Sometimes	no		2 no		1	1 no	Public_Transportation	Normal_Weight
Viale	26	1.85	105	yes	yes	3		3 Frequently	no		3 no		2	2 Sometimes	Public_Transportation	Obesity_Type_I
-emale	21	1.72	80	yes	yes	2		3 Frequently	no		2 yes		2	1 Sometimes	Public_Transportation	Overweight_Level_II

After Preprocessing



Model Development

Building a regression model;

Building And Fitting Regression model

```
In [13]: # Building Multiple Linear Regression model
    from sklearn.linear_model import LinearRegression
    model = LinearRegression()

In [14]: # Fitting Multiple Linear Regression to the Training set
    model.fit (X_train_scaled, y_train)

Out[14]: LinearRegression()
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook. On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

Model Evaluation

